

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Horvitz, H. Robert
Yuan, Junying
Shaham, Shai

(ii) TITLE OF INVENTION: CLONING, SEQUENCING AND CHARACTERIZATION
OF TWO CELL DEATH GENES AND USES THEREFOR

(iii) NUMBER OF SEQUENCES: 28

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Fish & Richardson P.C.
(B) STREET: 225 Franklin Street
(C) CITY: Boston
(D) STATE: MA
(E) COUNTRY: USA
(F) ZIP: 02110-2804

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/287,669
(B) FILING DATE: 09-AUG-1994
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 07/979,638
(B) FILING DATE: 20-NOV-1992

(viii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 07/897,788
(B) FILING DATE: 12-JUN-1992

(ix) ATTORNEY/AGENT INFORMATION:

(A) NAME: Clark, Paul T.
(B) REGISTRATION NUMBER: 30,162
(C) REFERENCE/DOCKET NUMBER: 01977/198005

(x) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 617/542-5070
(B) TELEFAX: 617/542-8906
(C) TELEX: 200154

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4407 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAATTCGCGT CGAATCATTG TCTGTTCCGT ATCGATTGAG AAACCGAAAC TTGTGATCGA

60

TAACAAGTCA TTCAAACACG GCGAAGATGT CTATGCGTAT AACAGAATAT TTGGAGAAAT	120
GCTCGCAAAA CTCGAAATTG TCACCGATAA AATGATTAAC TTGAAGGGGC TAATGTAAGT	180
TATCTGATGT TTCTACAATT AAAAAAATTG TTTTTTTTTC CAAATTAATT TTCGAAGATT	240
AACGAAAAAC GATTAAAAAT CAATAAAACG CAATAAAGAG GGCTTGGCTT TCTTTTTAAT	300
TTAAATTATA ATTTTTCTGA TTGTTGTATG AAGCTACAAA ATGTACTGTT TTTGTATTTG	360
AATATTGTAT TACAGGGTTG GGATTCTCGG CAAATATCAG CGACAGTGGG AGATTTAGAA	420
GAAGGACGTG TGACAATCAC TAAGTCAAAG AGGGAAAGGA TAAAGGATTG TGATATTTCA	480
CTGTTTTACT CATTGCTTTT TTAAATAAGA ACTATATGCC GATTTGCCGA TATATTTTTG	540
TTTATTAGGC CTCTCACATT CCTGTACAAT GTTTCTACCA AATAAACTGC ATTTTTATCT	600
GAAAATTCGA ATTTATTTTT GTCTACTTTT TACTCGTTGC ATTCGAGATC AGCATATCTT	660
CCGGTCTATT TATATTCAAC GATTTTTATA AATTAGTACT CCTTCATGTT TAATTTCAAT	720
TTATCTGTAA GCTTTACTGT ATTTTTTTAA AATCTTTCTT GCTTCTATCT GATTATACAA	780
TGTTCTTTAC TCATTTTCAA GGTATTTTAA TGCCTCACAA TTTATGCACA TTTCGGGCTT	840
GGAGATTTAT CCTCTATATT ACATGCCTGT TTTTTTAAAG GATATAATGT TTAACAAATA	900
ATTTTTTATC AATGCTATTG TATATTCTCC AGCTAACCGT TGTTTCGAAA ACATCACCTA	960
GCATTTTAAA ATTCACAAAA TCTTGCTTCC TTATAATCAA GAAGATTTTT CAGATGCTCT	1020
GCGAAATCGA ATGCCGCGCT TTGAGCACGG CACACACGAG GCTCATCCAC GACTTTGAAC	1080
CACGTGACGC ATTGACTTAT TTAGAAGGCA AAAACATTTT CACAGAAGAT CATTCTGAAC	1140
TTATCAGTAA AATGTCAACT CGCCTCGAGA GGATCGCCAA TTTTCTTCGA ATCTATCGAC	1200
GTCAAGCTTC TGAAC TTGGA CCACTCATCG ACTTTTTCAA CTACAACAAT CAAAGTCACC	1260
TTGCTGATTT CCTCGAAGAC TACATCGATT TTGCGATAAA TGAGCCAGAT CTA CTTCGTC	1320
CAGTAGTGAT TGCTCCACAA TTTTCCCGAC AAATGCTCGA TAGGAAACTA TTGCTTGGA	1380
ATGTTCCAAA ACAAATGACA TGCTATATTC GAGAGTATCA CGTGGATCGA GTGATCAAAA	1440
AGCTCGACGA GATGTGTGAT TTAGGTGAGA AAACTGGAAG CTCTCGTGTT TATTATAATC	1500
TTGCTTAAAC TTCAGACTCC TTTTTTCTGT TTCTACACGG CCGAGCTGGA TCCGGAAAAT	1560
CAGTAATTGC ATCACAAGCT CTTTCGAAAT CTGACCAACT TATTGGAATG TGAGTGGTAT	1620
TATCTGAATC TACGGATCTT CATTCTATTA CAGAAATTAT GATTCAATCG TTTGGCTCAA	1680
AGATAGTGGA ACAGCTCCAA AATCTACATT CGATTTATTT ACGGATATTT TGCTGATGCT	1740
AAAGTGAGTG AATAGAGTGC ATGTAACATT CAGCATGATT TTGAAATTAT GAAAATTTGA	1800
CCTGGTTAGC TTTTAATTTG ATATTTCGTG ACGCTTGCAT GTTTTGTGTG TTTGAAGACG	1860
AGCCCGTGTT GTGAGCGACA CGGATGACTC GCATTGATC ACCGACTTCA TTAACCGTGT	1920
TCTTTCAAGA AGCGAAGACG ATCTTCTCAA TTTCCCATCG GTGGAGCATG TCACGTCAGT	1980

TGTA	CTCAAA	AGGATGGTAA	GTTGCTTGCC	GATTCTGGTA	CAATATCTTA	AATTATTGGT	2040
TTTTAGATCT	GCAACGCACT	CATTGATCGT	CCAAATACTT	TATTCGTATT	TGATGACGTA		2100
GTTCAAGAAG	AAACAATTCG	TTGGGCTCAG	GAGCTACGTC	TTCGATGTCT	TGTA	ACTACT	2160
CGTGACGTGG	AAATATCAAA	TGCTGCTTCT	CAAACATGCG	AATTCATTGA	AGTGACATCA		2220
TTGGAAATCG	ATGAATGTTA	TGATTTTCTA	GAGCTTATG	GAATGCCGAT	GCCTGTTGGA		2280
GAAAAAGAAG	AAGATGTGCT	TAATAAAACA	ATCGAACTAA	GCAGTGGAAA	TCCAGCAACG		2340
CTTATGATGT	TTTTCAAGTC	TTGTGAACCG	AAAACATTTG	AAAAGTGAGT	GGGACATACC		2400
AATTTGAGAC	TTTTAAAATA	ATTTATTCTA	CAATAAAAGT	TAATCAAAAA	GTTTCATAGC		2460
TGATTGTCTT	TAAATTTTAC	GAATTGAGGA	TCAAAATCAA	GAATTAGGAT	CCTGGCACGA		2520
GAGAAACTG	TGTAGCTACC	GTACCCGAGA	GATTTTCTTG	ATATTTGCCA	TCGATTTAAT		2580
TTTTTAAGAA	AATTATCGTT	TTACATAATT	GAACAAGAGA	TACACGGTCT	CGACCCGACG		2640
GAAATTTTTT	AAATGAAAGC	GAGTATGAGC	CTGTTTTTCAT	TATTTTTTCGA	TTTTCTCTTG		2700
TTGTTTCTTT	TTATTTAAAG	CCTTTTATTT	TGAAACAAGT	CTAAAAATAT	TAAAAACTGA		2760
ATAAAATATT	TAAAAAAAT	CAAGTAAAT	AGAAAAACAG	CAAGGCTGGA	GA	CTACTGTA	2820
CTTCTTAAAT	CCGCATACTC	TTTTTATTTA	ATCATTTTCC	GGAATGTCGA	AACGAAATAA		2880
TACATTTTTTA	GTCCAAAATC	GCTAGGTATA	TTCTTAAAT	TATCAAACAT	TTTGCATTCA		2940
GAATGGCACA	GCTTAATAAC	AAATTGGAAA	GTCGAGGATT	AGTCGGTGTT	GAATGTATCA		3000
CCCCTTACTC	GTACAAGTCA	CTCGCAATGG	CTCTTCAAAG	ATGTGTTGAA	GTTTTGTCAG		3060
ATGAGGATCG	AAGTGCTCTT	GCTTTCGCAG	TTGTGATGCC	TCCTGGAGTT	GATATACCCG		3120
TCAAGCTATG	GTCATGTGTT	ATTCCAGTTG	ATATTTGTTC	AAATGAAGAA	GAACAATTGG		3180
ATGATGAAGT	TGCGGATCGG	TTGAAAAGAC	TCAGCAAGTA	TGAGTCTTGA	AATTTGAAGA		3240
TTTAAATTAA	CACCTTAAAT	TTCAGACGTG	GAGCTCTTCT	CAGTGGAAAA	CGAATGCCCG		3300
TTTTGACATT	CAAAATTGAT	CATATTATCC	ATATGTTCTT	GAAACACGTC	GTTGATGCAC		3360
AAACTATCGC	CGTATGCTGA	AAATGTCTCA	ACTTTCAATT	AAATTTTAAA	TTTTCAGAAT		3420
GGAATCTCAA	TTCTCGAGCA	GCGTCTTCTT	GAAATAGGAA	ACAATAATGT	ATCAGTACCG		3480
GAGCGACATA	TACCATCACA	TTTCCAAAAA	TTCCGTCGTT	CATCAGCCAG	TGAGATGTAT		3540
CCAAAACTA	CAGAAGAAAC	TGTGATCCGT	CCTGAAGACT	TCCCAAAGTT	CATGCAATTG		3600
CACCAGAAAT	TCTATGACTC	CCTCAAAAT	TTTGCATGCT	GTTAAACCT	ATCGTGTACA		3660
ATATTGCCTG	TATATTCCCC	TCGAAATACG	TTTATACTTT	TTCGCACGAG	TTTTCTCATT		3720
TTTTCATTTG	TACTTGTTTT	ATTTCTCTCC	AAAATTTTCA	ATCTATCCCA	AATGTTCTTA		3780
AATTTAATGT	TTTCTACAGA	TACTCAACAC	ATCTTGTTTC	ATCTCATCCT	TGCTTTTTTT		3840
TTTCAAATAT	ATTCAGTTTC	TTTTATAATT	TTAATTAATC	GAATTAATAC	ATTCACGTAA		3900

AGAATTCGCT GGACTATTAT TTTATCGCAT CCAAATGATT TATTCCCTAT TGTTCGAAAC 3960
 TTCCAAATTG ATCATTTTTA AACACGCCTC ATTAAATTGA AAGTCGTACT TTTAGTCTCG 4020
 AACATGAAGT AAGTTATTTT CTGTGTTCTA AATTCAAAGT GCATTCCAAA AGGACATTTG 4080
 ATGAGTTTTC ACGAAAACCG TAATTTTAC AATTTCCTTT CAGTTTGTAA GATGTTTCGAT 4140
 TTCTTTCCTC TGTTGGCGTC ATTACTACAT TTGCTTTGCT GCTTCACTTT ATCGAGATTC 4200
 TTGCCATCAA TGGAGTTCCA TCTAGACCGA TAGCAGTCTT CATATCATT TCCCTGTATA 4260
 TTGTACTGTT TCAGTATTTT AACTTATCGA TTACGTACTA TATTCAGTGG TTCACTGTTT 4320
 TCGGTCAATG GGTGACACGT GCTCGACGAN NAATTTTCAA CGAACGCAAT CTCCTAGTCA 4380
 CTTATCAACC AAGAGCCCTC ACCCATG 4407

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 549 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Leu Cys Glu Ile Glu Cys Arg Ala Leu Ser Thr Ala His Thr Arg
 1 5 10 15
 Leu Ile His Asp Phe Glu Pro Arg Asp Ala Leu Thr Tyr Leu Glu Gly
 20 25 30
 Lys Asn Ile Phe Thr Glu Asp His Ser Glu Leu Ile Ser Lys Met Ser
 35 40 45
 Thr Arg Leu Glu Arg Ile Ala Asn Phe Leu Arg Ile Tyr Arg Arg Gln
 50 55 60
 Ala Ser Glu Leu Gly Pro Leu Ile Asp Phe Phe Asn Tyr Asn Asn Gln
 65 70 75 80
 Ser His Leu Ala Asp Phe Leu Glu Asp Tyr Ile Asp Phe Ala Ile Asn
 85 90 95
 Glu Pro Asp Leu Leu Arg Pro Val Val Ile Ala Pro Gln Phe Ser Arg
 100 105 110
 Gln Met Leu Asp Arg Lys Leu Leu Leu Gly Asn Val Pro Lys Gln Met
 115 120 125
 Thr Cys Tyr Ile Arg Glu Tyr His Val Asp Arg Val Ile Lys Lys Leu
 130 135 140
 Asp Glu Met Cys Asp Leu Asp Ser Phe Phe Leu Phe Leu His Gly Arg
 145 150 155 160
 Ala Gly Ser Gly Lys Ser Val Ile Ala Ser Gln Ala Leu Ser Lys Ser
 165 170 175
 Asp Gln Leu Ile Gly Ile Asn Tyr Asp Ser Ile Val Trp Leu Lys Asp

180										185					190						
Ser	Gly	Thr	Ala	Pro	Lys	Ser	Thr	Phe	Asp	Leu	Phe	Thr	Asp	Ile	Leu						
		195					200					205									
Leu	Met	Leu	Lys	Ser	Glu	Asp	Asp	Leu	Leu	Asn	Phe	Pro	Ser	Val	Glu						
	210					215					220										
His	Val	Thr	Ser	Val	Val	Leu	Lys	Arg	Met	Ile	Cys	Asn	Ala	Leu	Ile						
	225				230					235					240						
Asp	Arg	Pro	Asn	Thr	Leu	Phe	Val	Phe	Asp	Asp	Val	Val	Gln	Glu	Glu						
				245					250					255							
Thr	Ile	Arg	Trp	Ala	Gln	Glu	Leu	Arg	Leu	Arg	Cys	Leu	Val	Thr	Thr						
			260					265					270								
Arg	Asp	Val	Glu	Ile	Ser	Asn	Ala	Ala	Ser	Gln	Thr	Cys	Glu	Phe	Ile						
		275					280					285									
Glu	Val	Thr	Ser	Leu	Glu	Ile	Asp	Glu	Cys	Tyr	Asp	Phe	Leu	Glu	Ala						
	290					295					300										
Tyr	Gly	Met	Pro	Met	Pro	Val	Gly	Glu	Lys	Glu	Glu	Asp	Val	Leu	Asn						
	305				310					315					320						
Lys	Thr	Ile	Glu	Leu	Ser	Ser	Gly	Asn	Pro	Ala	Thr	Leu	Met	Met	Phe						
				325					330					335							
Phe	Lys	Ser	Cys	Glu	Pro	Lys	Thr	Phe	Glu	Lys	Met	Ala	Gln	Leu	Asn						
			340					345					350								
Asn	Lys	Leu	Glu	Ser	Arg	Gly	Leu	Val	Gly	Val	Glu	Cys	Ile	Thr	Pro						
		355					360					365									
Tyr	Ser	Tyr	Lys	Ser	Leu	Ala	Met	Ala	Leu	Gln	Arg	Cys	Val	Glu	Val						
	370					375					380										
Leu	Ser	Asp	Glu	Asp	Arg	Ser	Ala	Leu	Ala	Phe	Ala	Val	Val	Met	Pro						
	385				390					395					400						
Pro	Gly	Val	Asp	Ile	Pro	Val	Lys	Leu	Trp	Ser	Cys	Val	Ile	Pro	Val						
				405					410					415							
Asp	Ile	Cys	Ser	Asn	Glu	Glu	Glu	Gln	Leu	Asp	Asp	Glu	Val	Ala	Asp						
			420					425					430								
Arg	Leu	Lys	Arg	Leu	Ser	Lys	Arg	Gly	Ala	Leu	Leu	Ser	Gly	Lys	Arg						
		435					440					445									
Met	Pro	Val	Leu	Thr	Phe	Lys	Ile	Asp	His	Ile	Ile	His	Met	Phe	Leu						
	4																				

Phe Pro Lys Phe Met Gln Leu His Gln Lys Phe Tyr Asp Ser Leu Lys
 530 535 540

Asn Phe Ala Cys Cys
 545

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Asp Gln Asp Lys Ser Gly Phe Ile Glu Glu Asp Glu
 1 5 10

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Asp Gln Asp Lys Asp Asp Phe Ile Gly Glu Asp Glu
 1 5 10

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Asp Ser Asp Gly Asp His Lys Ile Gly Val Asp Glu
 1 5 10

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Asp Ile Asn Lys Asp Asp Val Val Ser Trp Glu Glu

1

5

10

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Ala Lys Glu Gly Asp Pro Gln Leu Ser Lys Glu Glu
1 5 10

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Asp Lys Asn Gly Asp Gly Glu Cys Ser Phe Glu Glu
1 5 10

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Asp Ala Asp Gly Gly Gly Asp Ile Ser Val Lys Glu
1 5 10

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Asp Glu Asp Gly Ser Gly Thr Ile Asp Phe Glu Glu
1 5 10

(2) INFORMATION FOR SEQ ID NO:11:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: not relevant
 (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
- Asp Arg Asn Ala Asp Gly Tyr Ile Asp Ala Glu Glu
 1 5 10
- (2) INFORMATION FOR SEQ ID NO:12:
- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: not relevant
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
- Asp Lys Asn Asn Asp Gly Arg Ile Asp Phe Asp Glu
 1 5 10
- (2) INFORMATION FOR SEQ ID NO:13:
- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: not relevant
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
- Asp Lys Asp Gly Asn Gly Thr Ile Thr Thr Lys Glu
 1 5 10
- (2) INFORMATION FOR SEQ ID NO:14:
- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: not relevant
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
- Leu Gly Glu Asp Asn Ile Asn Val Val Glu Gly Asn Glu
 1 5 10
- (2) INFORMATION FOR SEQ ID NO:15:
- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 13 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: not relevant

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Asp Asn Asp Asn Asp Lys Phe Glu Gly Asn Cys Ala Glu
1 5 10

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Gly Val Asp Pro Ser Arg Lys Glu Asn His Leu Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Asp Leu Asn Lys Asp Gly Gln Ile Gln Ile Glu
1 5 10

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7653 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

AGATCTGAAA TAAGGTGATA AATTAATAAA TTAAGTGTAT TTCTGAGGAA ATTTGACTGT	60
TTTAGCACAA TTAATCTTGT TTCAGAAAAA AAGTCCAGTT TTCTAGATTT TTCCGTCTTA	120
TTGTCGAATT AATATCCCTA TTATCACTTT TTCATGCTCA TCCTCGAGCG GCACGTCCTC	180
AAAGAATTGT GAGAGCAAAC GCGCTCCCAT TGACCTCCAC ACTCAGCCGC CAAAACAAAC	240
GTTCGAACAT TCGTGTGTTG TGCTCCTTTT CCGTTATCTT GCAGTCATCT TTTGTCGTTT	300
TTTTCTTTGT TCTTTTTGTT GAACGTGTTG CTAAGCAATT ATTACATCAA TTGAAGAAAA	360
GGCTCGCCGA TTTATTGTTG CCAGAAAGAT TCTGAGATTC TCGAAGTCGA TTTTATAATA	420

TTTAACCTTG GTTTTTGCAT TGTTTCGTTT AAAAAACCA CTGTTTATGT GAAAAACGAT	480
TAGTTTACTA ATAAACTAC TTTTAAACCT TTACCTTTAC CTCACCGCTC CGTGTTTCATG	540
GCTCATAGAT TTTCGATACT CAAATCCAAA AATAAATTTA CGAGGGCAAT TAATGTGAAA	600
CAAAAACAAT CCTAAGATTT CCACATGTTT GACCTCTCCG GCACCTTCTT CCTTAGCCCC	660
ACCACTCCAT CACCTCTTTG GCGGTGTTCT TCGAAACCCA CTTAGGAAAG CAGTGTGTAT	720
CTCATTTGGT ATGCTCTTTT CGATTTTATA GCTCTTTGTC GCAATTTCAA TGCTTTAAAC	780
AATCCAAATC GCATTATATT TGTGCATGGA GGCAAATGAC GGGGTGGAA TCTTAGATGA	840
GATCAGGAGC TTTCAGGGTA AACGCCCCGT TCATTTTGTA CCACATTTCA TCATTTTCCT	900
GTCGTCCTTG GTATCCTCAA CTGTGCCCG TTTTGTTTTT GGTACACTCT TCCGTGATGC	960
CACCTGTCTC CGTCTCAATT ATCGTTTAGA AATGTGAACT GTCCAGATGG GTGACTCATA	1020
TTGCTGCTGC TACAATCCAC TTTCTTTTCT CATCGGCAGT CTTACGAGCC CATCATAAAC	1080
TTTTTTTTCC GCGAAATTTG CAATAAACCG GCCAAAAACT TTCTCCAAAT TGTTACGCAA	1140
TATATACAAT CCATAAGAAT ATCTTCTCAA TGTTTATGAT TTCTTCGCAG CACTTTCTCT	1200
TCGTGTGCTA ACATCTTATT TTTATAATAT TTCCGCTAAA ATTCCGATTT TTGAGTATTA	1260
ATTTATCGTA AAATTATCAT AATAGCACCG AAACTACTA AAAATGGTAA AAGCTCCTTT	1320
TAAATCGGCT CGACATTATC GTATTAAGGA ATCACAAAAT TCTGAGAATG CGTACTGCGC	1380
AACATATTTG ACGGCAAAAT ATCTCGTAGC GAAACTACA GTAATCTTT AAATGACTAC	1440
TGTAGCGCTT GTGTCGATTT ACGGGCTCAA TTTTGAAAA TAATTTTTTT TTTCGAATTT	1500
TGATAACCCG TAAATCGTCA CAACGCTACA GTAGTCATTT AAAGGATTAC TGTAGTTCTA	1560
GCTACGAGAT ATTTTGCGCG CCAAATATGA CTGTAATACG CATTCTCTGA ATTTTGTGTT	1620
TCCGTAATAA TTTCACAAGA TTTTGGCATT CCACTTTAAA GGCGCACAGG ATTTATTCCA	1680
ATGGGTCTCG GCACGCAAAA AGTTTGATAG ACTTTTAAAT TCTCCTTGCA TTTTAAATTC	1740
AATTACTAAA ATTTTCGTGA ATTTTCTGT TAAAATTTT AAAATCAGTT TTCTAATATT	1800
TTCCAGGCTG ACAAACAGAA ACAAAAACAC AACAAACATT TAAAAATCA GTTTTCAAAT	1860
TAAAAATAAC GATTTCTCAT TGAAAATTGT GTTTTATGTT TGCGAAAATA AAAGAGAACT	1920
GATTCAAAAC AATTTTAAAC AAAAAAACC CCAAATTCG CCAGAAATCA AGATAAAAAA	1980
TTCAAGAGGG TCAAAATTTT CCGATTTTAC TGACTTTCAC CTTTTTTTTT GTAGTTCAGT	2040
GCAGTTGTTG GAGTTTTTGA CGAAAAC TAGAAAAATC GATAAAAATT ACTCAAATCG	2100
AGCTGAATTT TGAGGACAAT GTTTAAAAA AAACACTATT TTTCCAATAA TTCACTCAT	2160
TTTCAGACTA AATCGAAAAT CAAATCGTAC TCTGACTACG GGTCAGTAGA GAGGTCAACC	2220
ATCAGCCGAA GATGATGCGT CAAGATAGAA GGAGCTTGCT AGAGAGGAAC ATTATGATGT	2280
TCTCTAGTCA TCTAAAAGTC GATGAAATTC TCGAAGTTCT CATCGCAAAA CAAGTGTGTA	2340

ATAGTGATAA TGGAGATATG ATTAATGTGA GTTTTTAATC GAATAATAAT TTTAAAAAAA	2400
AATTGATAAT ATAAAGAATA TTTTTCAGT CATGTGGAAC GGTTCGCGAG AAGAGACGGG	2460
AGATCGTGAA AGCAGTGCAA CGACGGGGAG ATGTGGCGTT CGACGCGTTT TATGATGCTC	2520
TTCGCTCTAC GGGACACGAA GGA CTGCTG AAGTTCTTGA ACCTCTCGCC AGATCGTAGG	2580
TTTTTAAAGT TCGGCGCAA AGCAAGGGTC TCACGGAAAA AAGAGGCGGA TCGTAATTTT	2640
GCAACCCACC GGCACGGTTT TTTCTCCGA AAATCGGAAA TTATGCACTT TCCCAAATAT	2700
TTGAAGTGAA ATATATTTTA TTTACTGAAA GCTCGAGTGA TTATTTATTT TTTAACACTA	2760
ATTTTCGTGG CGCAAAGGC CATTTTGTAG ATTTGCCGAA AATACTTGTC ACACACACAC	2820
ACACACATCT CCTTCAAATA TCCCTTTTTT CAGTGTGAC TCGAATGCTG TCGAATTCGA	2880
GTGTCCAATG TCACCGGCAA GCCATCGTCG GAGCCGCGCA TTGAGCCCCG CCGGCTACAC	2940
TTCACCGACC CGAGTTCACC GTGACAGCGT CTCTTCAGTG TCATCATTCA CTTCTTATCA	3000
GGATATCTAC TCAAGAGCAA GATCTCGTTC TCGATCGCGT GCACTTCATT CATCGGATCG	3060
ACACAATTAT TCATCTCCTC CAGTCAACGC ATTTCCCAGC CAACCTTGTA TGTGATGCG	3120
AACACTAAAT TCTGAGAATG CGCATTACTC AACATATTTG ACGCGCAAAT ATCTCGTAGC	3180
GAAAAATACA GTAACCCTTT AAATGACTAT TGTAGTGTG ATTTACGGGC TCGATTTTCG	3240
AAACGAATAT ATGCTCGAAT TGTGACAACG AATTTTAATT TGTCAATTTT GTGTTTTCTT	3300
TTGATATTTT TGATCAATTA ATAAATTATT TCCGTAAACA GACACCAGCG CTACAGTACT	3360
CTTTTAAAGA GTTACAGTAG TTTTCGCTTC AAGATATTTT GAAAAGAATT TTAAACATTT	3420
TGAAAAAATA TCATCTAACA TGTGCCAAAA CGCTTTTTTC AAGTTTCGCA GATTTTTTGA	3480
TTTTTTTCAT TCAAGATATG CTTATTAACA CATATAATTA TCATTAATGT GAATTTCTTG	3540
TAGAAATTTT GGGCTTTTCG TTCTAGTATG CTCTACTTTT GAAATTGCTC AACGAAAAAA	3600
TCATGTGGTT TGTTCATATG AATGACGAAA AATAGCAATT TTTTATATAT TTTCCCCTAT	3660
TCATGTTGTG CAGAAAAATA GTAAAAAAGC GCATGCATTT TTCGACATTT TTTACATCGA	3720
ACGACAGCTC ACTTCACATG CTGAAGACGA GAGACGCGGA GAAATACCAC ACATCTTTCT	3780
GCGTCTCTCG TCTTCAGCAT GTGAAATGGG ATCTCGGTCG ATGTAAAAAA ATGTGGAATA	3840
ATGTAAAAAA TGCATGCGTT TTTTACACT TTTCTGCACA AATGAATAGG GGGAAAATGT	3900
ATTAAATAC ATTTTTTGTA TTTTCAACA TCACATGATT AACCCCATTA TTTTTCGTT	3960
GAGCAACTTA AAAAGTAGAG AATATTAGAG CGAAAACCAA AATTTCTTCA AGATATTACC	4020
TTTATTGATA ATTATAGATG TTAATAAGCA TATCTTGAAT GAAAGTCAGC AAAAATATGT	4080
GCGAAACACC TGAAAAAAT CAAAAATTCT GCGAAAATTG AAAAAATGCA TTAAATACA	4140
TTTTTGCAAT TTTCTACATC ACATGAATGT AGAAAATTAA AAGGGAAATC AAAATTTCTA	4200
GAGGATATAA TTGAATGAAA CATTGCGAAA TTAAATGTG CGAAACGTCA AAAAAGAGGA	4260

AATTTGGGTA TCAAAATCGA TCCTAAAACC AACACATTTC AGCATCCGCC AACTCTTCAT	4320
TCACCGGATG CTCTTCTCTC GGATACAGTT CAAGTCGTAA TCGCTCATT AGCAAAGCTT	4380
CTGGACCAAC TCAATACATA TTCCATGAAG AGGATATGAA CTTTGTCGAT GCACCAACCA	4440
TAAGCCGTGT TTTGACGAG AAAACCATGT ACAGAACTT CTCGAGTCCT CGTGGAATGT	4500
GCCTCATCAT AAATAATGAA CACTTTGAGC AGATGCCAAC ACGGAATGGT ACCAAGGCCG	4560
ACAAGGACAA TCTTACCAAT TTGTTTCAGAT GCATGGGCTA TACGGTTATT TGCAAGGACA	4620
ATCTGACGGG AAGGGTACGG CGAAATTATA TTACCCAAAC GCGAAATTG CCATTTTGCG	4680
CCGAAAATGT GCGCCCCGT CTCGACACGA CAATTTGTGT TAAATGCAA AATGTATAAT	4740
TTTGCAAAA ACAAATTTT GAACTTCCGC GAAAATGATT TACCTAGTTT CGAAATTTTC	4800
GTTTTTCCG GCTACATTAT GTGTTTTTTC TTAGTTTTTC TATAATATTT GATGTAAAA	4860
ACCGTTTGTA AATTTTCAGA CAATTTTCCG CATACAAAC TTGATAGCAC GAAATCAATT	4920
TTCTGAATTT TCAAAATTAT CCAAAAATGC ACAATTTAAA ATTTGTGAAA ATTGGCAAAC	4980
GGTGTTCOA TATGAAATGT ATTTTAAAA ACTTTAAAA CCACTCCGGA AAAGCAATAA	5040
AAATCAAAAC AACGTCACAA TTCAAATTCA AAAGTTATTC ATCCGATTG TTTATTTTGT	5100
CAAAATTTGA AAAAATCATG AAGGATTTAG AAAAGTTTTA TAACATTTT TCTAGATTTT	5160
TCAAAATTTT TTTTAACAAA TCGAGAAAAA GAGAATGAAA AATCGATTTT AAAAATATCC	5220
ACAGCTTCGA GAGTTTGAAA TTACAGTACT CCTTAAAGGC GCACACCCCA TTTGCATTGG	5280
ACCAAAAATT TGTCGTGTCG AGACCAGGTA CCGTAGTTTT TGTCGCAAAA ATTGCACCAT	5340
TGGACAATAA ACCTTCCTAA TCACCAAAAA GTAAAATTGA AATCTTCGAA AAGCCAAAAA	5400
ATTCAAAAA AAAGTCGAAT TTCGATTTTT TTTTGGTTT TTTGGTCCCA AAAACCAAAA	5460
AAATCAATTT TCTGCAAAAT ACCAAAAAGA AACCCGAAAA AATTTCCCAG CCTTGTTCTT	5520
AATGTAACT GATATTTAAT TTCCAGGGAA TGCTCCTGAC AATTCGAGAC TTTGCCAAAC	5580
ACGAATCACA CGGAGATTCT GCGATACTCG TGATTCTATC ACACGGAGAA GAGAATGTGA	5640
TTATTGGAGT TGATGATATA CCGATTAGTA CACACGAGAT ATATGATCTT CTCAACGCGG	5700
CAAATGCTCC CCGTCTGGCG AATAAGCCGA AAATCGTTTT TGTGCAGGCT TGTGAGGCG	5760
GTTCTTTTTT TATTTTAATT TTAATATAA TATTTTAAAT AAATTCATTT TCAGAACGTC	5820
GTGACAATGG ATTCCCAGTC TTGGATTCTG TCGACGGAGT TCCTGCATTT CTTCTGCTG	5880
GATGGGACAA TCGAGACGGG CCATTGTTCA ATTTTCTTGG ATGTGTGCGG CCGCAAGTTC	5940
AGGTTGCAAT TTAATTTCTT GAATGAGAAT ATTCCTTCAA AAAATCTAAA ATAGATTTTT	6000
ATTCCAGAAA GTCCCGATCG AAAAATTGCG ATATAATTAC GAAATTTGTG ATAAAATGAC	6060
AAACCAATCA GCATCGTCGA TCTCCGCCA CTTATCGGA TTGGTTTGAA AGTGGGCGGA	6120
GTGAATTGCT GATTGGTCGC AGTTTTCACT TTAGAGGGAA TTAAAAATC GCCTTTTCGA	6180

AAATTAAAAA TTGATTTTTT CAATTTTTTC GAAAAATATT CCGATTATTT TATATTCTTT	6240
GGAGCGAAAG CCCCGTCCTG TAAACATTTT TAAATGATAA TTAATAAATT TTTGCAGCAA	6300
GTGTGGAGAA AGAAGCCGAG CCAAGCTGAC ATTCTGATTC GATACGCAAC GACAGCTCAA	6360
TATGTTTCGT GGAGAAACAG TGCTCGTGGA TCATGGTTCA TTCAAGCCGT CTGTGAAGTG	6420
TTCTCGACAC ACGCAAAGGA TATGGATGTT GTTGAGCTGC TGAAGAAGT CAATAAGAAG	6480
GTCGCTTGTG GATTTTCAGAC ATCACAGGGA TCGAATATTT TGAAACAGAT GCCAGAGGTA	6540
CTTGAAACAA ACAATGCATG TCTAACTTTT AAGGACACAG AAAAATAGGC AGAGGCTCCT	6600
TTTGCAAGCC TGCCGCGCGT CAACCTAGAA TTTTAGTTTT TAGCTAAAAT GATTGATTTT	6660
GAATATTTTA TGCTAATTTT TTTGCGTTAA ATTTTGAAAT AGTCACTATT TATCGGGTTT	6720
CCAGTAAAAA ATGTTTATTA GCCATTGGAT TTTACTGAAA ACGAAAATTT GTAGTTTTTC	6780
AACGAAATTT ATCGATTTTT AAATGTAAAA AAAAATAGCG AAAATTACAT CAACCATCAA	6840
GCATTTAAGC CAAAATTGTT AACTCATTTA AAAATTAATT CAAAGTTGTC CACGAGTATT	6900
ACACGGTTGG CGCGCGGCAA GTTTGCAAAA CGACGCTCCG CCTCTTTTTT TGTGCGGCTT	6960
GAAAACAAGG GATCGGTTTA GATTTTTCCC CAAAATTTAA ATTAAATTC AGATGACATC	7020
CCGCCTGCTC AAAAAGTTCT ACTTTTGCC GGAAGCACGA AACTCTGCCG TCTAAAATTC	7080
ACTCGTGATT CATTGCCCAA TTGATAATTG TCTGTATCTT CTCCCCAGT TCTCTTTCGC	7140
CCAATTAGTT TAAAACCATG TGTATATTGT TATCCTATAC TCATTTCACT TTATCATTCT	7200
ATCATTTCTC TTCCCATTTT CACACATTTT CATTTCTCTA CGATAATCTA AAATTATGAC	7260
GTGTGTGTCT CGAACGCATA ATAATTTTAA TAACTCGTTT TGAATTTGAT TAGTTGTTGT	7320
GCCCAGTATA TATGTATGTA CTATGCTTCT ATCAACAAAA TAGTTTCATA GATCATCACC	7380
CCAACCCAC CAACCTACCG TACCATATTC ATTTTGGCCG GGAATCAATT TCGATTAATT	7440
TTAACCTATT TTTTCGCCAC AAAAAATCTA ATATTTGAAT TAACGAATAG CATTCCCATC	7500
TCTCCCGTGC CGGAATGCCT CCCGGCCTTT TAAAGTTCGG AACATTTGGC AATTATGTAT	7560
AAATTTGTAG GTCCCCCCCA TCATTTCCCG CCCATCATCT CAAATTGCAT TCTTTTTTCG	7620
CCGTGATATC CCGATTCTGG TCAGCAAAGA TCT	7653

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 503 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: not relevant
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met	Met	Arg	Gln	Asp	Arg	Arg	Ser	Leu	Leu	Glu	Arg	Asn	Ile	Met	Met
1				5					10					15	

Phe Ser Ser His Leu Lys Val Asp Glu Ile Leu Glu Val Leu Ile Ala
20 25 30
Lys Gln Val Leu Asn Ser Asp Asn Gly Asp Met Ile Asn Ser Cys Gly
35 40 45
Thr Val Arg Glu Lys Arg Arg Glu Ile Val Lys Ala Val Gln Arg Arg
50 55 60
Gly Asp Val Ala Phe Asp Ala Phe Tyr Asp Ala Leu Arg Ser Thr Gly
65 70 75 80
His Glu Gly Leu Ala Glu Val Leu Glu Pro Leu Ala Arg Ser Val Asp
85 90 95
Ser Asn Ala Val Glu Phe Glu Cys Pro Met Ser Pro Ala Ser His Arg
100 105 110
Arg Ser Arg Ala Leu Ser Pro Ala Gly Tyr Thr Ser Pro Thr Arg Val
115 120 125
His Arg Asp Ser Val Ser Ser Val Ser Ser Phe Thr Ser Tyr Gln Asp
130 135 140
Ile Tyr Ser Arg Ala Arg Ser Arg Ser Arg Ser Arg Ala Leu His Ser
145 150 155 160
Ser Asp Arg His Asn Tyr Ser Ser Pro Pro Val Asn Ala Phe Pro Ser
165 170 175
Gln Pro Ser Ser Ala Asn Ser Ser Phe Thr Gly Cys Ser Ser Leu Gly
180 185 190
Tyr Ser Ser Ser Arg Asn Arg Ser Phe Ser Lys Ala Ser Gly Pro Thr
195 200 205
Gln Tyr Ile Phe His Glu Glu Asp Met Asn Phe Val Asp Ala Pro Thr
210 215 220
Ile Ser Arg Val Phe Asp Glu Lys Thr Met Tyr Arg Asn Phe Ser Ser
225 230 235 240
Pro Arg Gly Met Cys Leu Ile Ile Asn Asn Glu His Phe Glu Gln Met
245 250 255
Pro Thr Arg Asn Gly Thr Lys Ala Asp Lys Asp Asn Leu Thr Asn Leu
260 265 270
Phe Arg Cys Met Gly Tyr Thr Val Ile Cys Lys Asp Asn Leu Thr Gly
275 280 285
Arg Gly Met Leu Leu Thr Ile Arg Asp Phe Ala Lys His Glu Ser His
290 295 300
Gly Asp Ser Ala Ile Leu Val Ile Leu Ser His Gly Glu Glu Asn Val
305 310 315 320
Ile Ile Gly Val Asp Asp Ile Pro Ile Ser Thr His Glu Ile Tyr Asp
325 330 335
Leu Leu Asn Ala Ala Asn Ala Pro Arg Leu Ala Asn Lys Pro Lys Ile
340 345 350
Val Phe Val Gln Ala Cys Arg Gly Glu Arg Arg Asp Asn Gly Phe Pro

355	360	365
Val Leu Asp Ser Val Asp Gly	Val Pro Ala Phe Leu Arg Arg Gly Trp	
370	375	380
Asp Asn Arg Asp Gly Pro Leu Phe Asn Phe Leu Gly Cys Val Arg Pro		
385	390	395
Gln Val Gln Gln Val Trp Arg Lys Lys Pro Ser Gln Ala Asp Ile Leu		
405	410	415
Ile Arg Tyr Ala Thr Thr Ala Gln Tyr Val Ser Trp Arg Asn Ser Ala		
420	425	430
Arg Gly Ser Trp Phe Ile Gln Ala Val Cys Glu Val Phe Ser Thr His		
435	440	445
Ala Lys Asp Met Asp Val Val Glu Leu Leu Thr Glu Val Asn Lys Lys		
450	455	460
Val Ala Cys Gly Phe Gln Thr Ser Gln Gly Ser Asn Ile Leu Lys Gln		
465	470	475
Met Pro Glu Met Thr Ser Arg Leu Leu Lys Lys Phe Tyr Phe Trp Pro		
485	490	495
Glu Ala Arg Asn Ser Ala Val		
500		

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 505 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Met Arg Gln Asp Arg Trp Leu Leu Glu Arg Asn Ile Leu Glu Phe	
1	15
Ser Ser Lys Leu Gln Ala Asp Leu Ile Leu Asp Val Leu Ile Ala Lys	
20	30
Gln Val Leu Asn Ser Asp Asn Gly Asp Val Ile Asn Ser Cys Arg Thr	
35	45
Glu Arg Asp Asn Glu Lys Glu Ile Val Lys Ala Val Gln Arg Arg Gly	
50	60
Asp Glu Ala Phe Asp Ala Phe Tyr Asp Ala Leu Arg Asp Thr Gly His	
65	80
Asn Asp Leu Ala Asp Val Leu Met Pro Leu Ser Arg Pro Xaa Xaa Xaa	
85	95
Asn Pro Val Pro Met Glu Cys Pro Met Ser Pro Ser Ser His Arg Arg	
100	110
Ser Arg Ala Leu Ser Pro Pro Xaa Tyr Ala Ser Pro Thr Arg Val His	
115	125

Arg Asp Ser Ile Ser Ser Val Ser Ser Phe Thr Ser Thr Tyr Gln Asp
 130 135 140
 Val Tyr Ser Arg Ala Arg Ser Ser Ser Arg Ser Ser Arg Pro Leu Gln
 145 150 155 160
 Ser Ser Asp Arg His Asn Tyr Met Ser Ala Ala Thr Ser Phe Pro Ser
 165 170 175
 Gln Pro Xaa Ser Ala Asn Ser Ser Phe Thr Gly Cys Ala Ser Leu Gly
 180 185 190
 Tyr Ser Ser Ser Arg Asn Arg Ser Phe Ser Lys Thr Ser Ala Gln Ser
 195 200 205
 Gln Tyr Ile Phe His Glu Glu Asp Met Asn Tyr Val Asp Ala Pro Thr
 210 215 220
 Ile His Arg Val Phe Asp Glu Lys Thr Met Tyr Arg Asn Phe Ser Ser
 225 230 235 240
 Pro Arg Gly Leu Cys Leu Ile Ile Asn Asn Glu His Phe Glu Gln Met
 245 250 255
 Pro Thr Arg Asn Gly Thr Lys Ala Asp Lys Asp Asn Leu Thr Asn Ile
 260 265 270
 Phe Arg Cys Met Gly Tyr Thr Val Ile Cys Lys Asp Asn Leu Thr Gly
 275 280 285
 Arg Glu Met Leu Ser Thr Ile Arg Ser Phe Gly Arg Asn Asp Met His
 290 295 300
 Gly Asp Ser Ala Ile Leu Val Ile Leu Ser His Gly Glu Xaa Asn Val
 305 310 315 320
 Ile Ile Gly Val Asp Asp Val Ser Val Asn Val His Glu Ile Tyr Asp
 325 330 335
 Leu Leu Asn Ala Ala Asn Ala Pro Arg Leu Ala Asn Lys Pro Lys Leu
 340 345 350
 Val Phe Val Gln Ala Cys Arg Gly Glu Arg Arg Asp Asn Gly Phe Pro
 355 360 365
 Val Leu Asp Ser Val Asp Gly Val Pro Ser Leu Ile Arg Arg Gly Trp
 370 375 380
 Asp Asn Arg Asp Gly Pro Leu Phe Asn Phe Leu Gly Cys Val Arg Pro
 385 390 395 400
 Gln Val Gln Gln Val Trp Arg Lys Lys Pro Ser Gln Ala Asp Met Leu
 405 410 415
 Ile Ala Tyr Ala Thr Thr Ala Gln Tyr Val Ser Trp Arg Asn Ser Ala
 420 425 430
 Arg Gly Ser Trp Phe Ile Gln Ala Val Cys Glu Val Phe Ser Leu His
 435 440 445
 Ala Lys Asp Met Asp Val Val Glu Leu Leu Thr Glu Val Asn Lys Lys
 450 455 460
 Val Ala Cys Gly Phe Gln Thr Ser Gln Gly Ser Asn Ile Leu Lys Gln

465

470

475

480

Met Pro Glu Leu Thr Ser Arg Leu Leu Lys Lys Phe Tyr Phe Trp Pro
 485 490 495

Glu Asp Arg Gly Arg Asn Ser Ala Val
 500 505

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 479 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: not relevant
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Thr Val Ser Ile Ser Leu Ile Ile Ala Arg Gln Val Leu Asn Ser Asp
 1 5 10 15
 Asn Gly Asp Met Ile Asn Ser Cys Gly Thr Val Arg Glu Lys Arg Arg
 20 25 30
 Glu Ile Val Lys Ala Val Gln Arg Arg Gly Asp Val Ala Phe Asp Ala
 35 40 45
 Phe Tyr Asp Ala Leu Arg Ser Thr Gly His Glu Gly Leu Ala Glu Val
 50 55 60
 Leu Glu Pro Leu Ala Arg Ser Val Asp Ser Asn Ala Val Glu Phe Glu
 65 70 75 80
 Cys Pro Met Ser Pro Ala Ser His Arg Arg Ser Arg Ala Leu Ser Pro
 85 90 95
 Ala Gly Tyr Thr Ser Pro Thr Arg Val His Arg Asp Ser Val Ser Ser
 100 105 110
 Val Ser Ser Phe Thr Ser Tyr Gln Asp Ile Tyr Ser Arg Ser Thr Ser
 115 120 125
 Ser Ser Ser Arg Pro Leu His Thr Ser Asp Arg His Asn Tyr Val Ser
 130 135 140
 Pro Ser Asn Ser Phe Gln Ser Gln Pro Ala Ser Ala Asn Ser Ser Phe
 145 150 155 160
 Thr Gly Ser Ser Ser Leu Gly Tyr Ser Ser Ser Arg Thr Arg Ser Tyr
 165 170 175
 Ser Lys Ala Ser Ala His Ser Gln Tyr Ile His Glu Glu Asp Met Asn
 180 185 190
 Tyr Val Asp Ala Pro Thr Ile His Arg Val Phe Asp Glu Lys Thr Met
 195 200 205
 Tyr Arg Asn Phe Ser Thr Pro Arg Gly Leu Cys Leu Ile Ile Asn Asn
 210 215 220
 Glu His Phe Glu Gln Met Pro Thr Arg Asn Gly Thr Lys Pro Asp Lys
 225 230 235 240

Asp Asn Ile Ser Asn Leu Phe Arg Cys Met Gly Tyr Ile Val His Cys
 245 250 255
 Lys Asp Asn Leu Thr Gly Arg Gly Met Met Leu Thr Ile Arg Asp Phe
 260 265 270
 Ala Lys Asn Glu Thr His Gly Asp Ser Ala Ile Leu Val Ile Leu Ser
 275 280 285
 His Gly Glu Glu Asn Val Ile Ile Gly Val Asp Asp Val Ser Val Asn
 290 295 300
 Val His Glu Ile Tyr Xaa Leu Leu Asn Ala Ala Asn Ala Pro Arg Leu
 305 310 315 320
 Ala Asn Lys Pro Lys Leu Val Phe Val Gln Ala Cys Arg Gly Glu Arg
 325 330 335
 Arg Asp Val Gly Phe Pro Val Leu Asp Ser Val Asp Gly Val Pro Ala
 340 345 350
 Leu Ile Arg Arg Gly Trp Asp Lys Gly Asp Gly Pro Leu Phe Asn Phe
 355 360 365
 Leu Gly Cys Val Arg Pro Gln Ala Gln Gln Val Trp Arg Lys Lys Pro
 370 375 380
 Ser Gln Ala Asp Ile Leu Ile Ala Tyr Ala Thr Thr Ala Gln Tyr Val
 385 390 395 400
 Ser Trp Arg Asn Ser Ala Arg Gly Ser Trp Phe Ile Gln Ala Val Cys
 405 410 415
 Glu Val Phe Ser Leu His Ala Lys Asp Met Asp Val Val Glu Leu Leu
 420 425 430
 Thr Glu Val Asn Lys Lys Val Ala Cys Gly Phe Gln Thr Ser Gln Gly
 435 440 445
 Ala Asn Ile Leu Lys Gln Met Pro Glu Leu Thr Ser Arg Leu Leu Lys
 450 455 460
 Lys Phe Tyr Phe Trp Pro Glu Asp Arg Asn Arg Ser Ser Ala Val
 465 470 475

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ATTGGCGATC CTCTCGA

17

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TCATCGACTT TTAGATGACT AGAGAACATC

30

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GTTGCACTGC TTTCACGATC TCCCGTCTCT

30

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GTTTAATTAC CCAAGTTTGA G

21

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GGTTTAAACC AGTTACTCAA G

21

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

CCGGTGACAT TGGACACTC

19

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ACTATTCAAC ACTTG

15